THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Partner selection into policy relevant field experiments

Citation for published version:
Belot, M & James, J 2013 'Partner selection into policy relevant field experiments' ESE Discussion Papers, no. 236, Edinburgh School of Economics Discussion Paper Series.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Publisher Rights Statement:
© Belot, M., & James, J. (2013). Partner Selection into Policy Relevant Field Experiments. Edinburgh School of Economics Discussion Paper Series.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 05. May. 2023

Edinburgh School of Economics

Discussion Paper Series

Number 236

Partner Selection into Policy Relevant Field Experiments

Michèle Belot
(University of Edinburgh)

Jonathan James
(University of Bath)

Date
December 2013

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://edin.ac/16ja6A6

THE UNIVERSITY *of* EDINBURGH

Partner Selection into Policy Relevant Field Experiments*

Michele Belot (University of Edinburgh)

Jonathan James (University of Bath)

December 17, 2013

Abstract

This study investigates the issue of self-selection of stakeholders into participation and collaboration in policy-relevant experiments. We document and test the implications of self-selection in the context of randomised policy experiment we conducted in primary schools in the UK. The main questions we ask are (1) is there evidence of selection on key observable characteristics likely to matter for the outcome of interest and (2) does selection matter for the estimates of treatment effects. The experimental work consists in testing the effects of an intervention aimed at encouraging children to make more healthy choices at lunch. We recruited schools through local authorities and randomised schools across two incentive treatments and a control group. We document the selection taking place both at the level of local authorities and at the school level. Overall we find mild evidence of selection on key observables such as obesity levels and socio-economic characteristics. We find evidence of selection along indicators of involvement in healthy lifestyle programmes at the school level, but the magnitude is small. Moreover, We do not find significant differences in the treatment effects of the experiment between variables which, albeit to a mild degree, are correlated with selection into the experiment. To our knowledge, this is the first study providing direct evidence on the magnitude of self-selection in field experiments.

JEL Classification: C93, I18, J13

*We thank seminar participants at the University of Bath, and Oxford, and conference participants at the European Economic Association conference (Malaga, 2012), Royal Economic Society conference (Royal Holloway, London, 2013), Florence Workshop on Behavioural and Experimental Economics (May 2013), International Conference on Public Policy (Grenoble, June 2013)

Keywords: Selection, Field Experiments, Randomised controlled trials, External Validity

1 Introduction

The last decade has seen a booming increase in the popularity of field experiments in economics and the social sciences (Holt, 2005). The main driving factors behind this increasing popularity are, on the one hand, the quest for identification of causal mechanisms – which is easier to achieve when researchers are directly involved in manipulating the economic environment of interest – and, on the other hand, a quest to remain close to reality as opposed to studying subjects in an isolated laboratory context. There is now a stronghold of researchers advocating the case for randomized controlled trials (RCTs) in social policy (Burtless, 1995, Duflo and Kremer, 2005).

This study is interested specifically in a fundamental methodological issue associated with field experimental research: the selection of field collaborators into the experiment. Conducting field experiments usually requires finding collaborators such as employers, policymakers, schools, etc. who are prepared to collaborate with researchers and provide the necessary support for data collection. As List (2011) puts, the support of a key person prepared to stand behind the research project is often critical: "*Have a champion within the organization – the higher up the better. Making the experiment a "we" project instead of an "us versus them" pursuit as early as possible is critical.*" Surprisingly perhaps, field experimental papers devote very little attention to the issue of "selection into the experiment". As an illustration, we provide in Appendix A a brief overview of the information provided in field experimental studies published over the last five years in the top 5 journals and in the American Economic Journal Applied. We focus on the fields of policy evaluation, personnel economics and development economics, which have all experienced a significant increase in the popularity of field experimental research.[1] In most cases we know little or nothing about how the collaborators were selected and approached, and the experimental sample is not compared to the broader population of interest. One notable exception is a recent paper by Fryer (2011).

Researchers are well aware of the limitations that result from restricting experiments to subjects who have opted in. The main limitation is a possible selection bias and a lack of external validity. Obviously there is always some form of selection taking place. An experiment will, for example, take place within a given geographical area and at a particular point in time. This initial selection is often for practical reasons. Researchers located in California will find it more practical to conduct a field experiment in California than across the entire United States or across several countries. It is probably even desirable that the experiment can be conducted with sufficient oversight by the researchers.

[1] We searched these journals systematically for the keywords 'field' and 'experiment'.

It does nevertheless raise issues of generalizability, insofar as it restricts the sample to a population with certain characteristics (e.g. people living in California at a particular time).

How important is selection in field experimental research? Harrison and List (2004) acknowledge in their review paper that we know very little about the implications of self-selection for field experimental research. Seven years later Ludwig et al. (2011) point out that this is still an open question that has not been answered.

We study the issue of self-selection in participation in a randomised controlled experiment conducted in a highly policy relevant domain: children's diet. The goal of the experiment is to test the effectiveness of various incentive schemes to encourage children to eat fruit and vegetables at lunch. To conduct this experiment, we sought the collaboration of primary schools in the UK. We recruited them through local authorities, which play an overarching and coordinating role in the UK. Rather than picking a set of local authorities in an arbitrary manner, we approached all local authorities in the country at the same time and in the same manner (via e-mail) and asked whether they would be interested in collaborating with us. If they responded positively, we asked them to bring us in contact with at least five local schools. the rationale for doing this is to understand how overarching bodies themselves select potential participants for randomised controlled experiments. Since many social experiments are piloted by government bodies, this type of selection is likely to be highly relevant in the design of many social experiments. The randomisation does eventually take place at the school level and within local authorities, so local authorities should expect some schools to be treated and some schools to be part of a control group. The schools are the ones that are ultimately directly involved in the experiment and data collection. We contacted the schools suggested by the local authorities, briefed them about the project and they then decided whether to participate or not. Thus, we have potential selection operating at different levels: self-selection of local authorities, selection of schools by local authorities and self-selection of schools into the experiment.

We document how selection operates at these differentlevels along observable characteristics of the population under consideration – characteristics that we would expect could matter in the decision to participate, such as obesity rates and socio-economic indicators. We consider a wide range of variables that could a priori be relevant and see whether they are correlated with selection or not. Then we investigate whether the variables that are correlated with selection are also correlated with treatment effects estimated in the RCT. Of course we can only do this exercise with observables, and cannot provide a full

account of the selection bias that could take place. It is clear that we cannot control for some key variables (such as personality characteristics of the people involved) that may introduce a selection bias in the RCT as well. Moreover, it is clear that if selection was only driven by observables, it would be straightforward to correct for it. The point here is to get a sense of how much selection takes place along characteristics that could a priori be relevant and are observable, which should in principle give a sense of the importance of the full selection problem.

We propose that the documentation exercise we perform here should be a minimum standard for field experimental research in social sciences. It seems obvious that collecting information on key variables at the early stages of an experiment can provide useful insights into the possible magnitude of selection and is helpful to see how the results can be generalized to the population at large, which is the overriding aim of pilot studies and policy-relevant experiments. At the moment, researchers tend to be very conservative and modest in their claims regarding the external validity of their results. We believe that a proper documentation of the selection process would help us drawing more general lessons from field randomised controlled experiments.

To our knowledge, this is the first study that documents systematically the process of selection into a policy-related experiment. We have the advantage of having access to a range of indicators on the broad population of interest (all local authorities in England and all primary schools within each local authority) and study systematically the correlation between these indicators and selection into collaborating with us.

Our findings can be summarised as follows. First, we find that local authorities who self-select tend to be larger and richer, and have less favourable characteristics in terms of the outcome of interest (e.g. lower rates of fruit and vegetable consumption). But the selection is mild and almost disappears when it comes to actual participation. We also find no evidence that selection operates according to pre-trends. We also find little evidence of selection at the school level, except for one dimension, which is that the schools suggested by the local authorities to conduct the experiment are more likely to be involved in programmes promoting healthy lifestyles (as evaluated by an independent official body). Second, we do not find any significant correlations between the treatment effects of the experiment and the variables which, albeit to a mild degree, are correlated with selection into the experiment.

The remain of the paper is structured as follows. Section 2 discusses related work. Section 3 presents the experimental design and the context of our study. In Section 4, we look at the determinants of selection at the local authority and school levels. In Section

5, we examine to what extent the treatment effects estimated in the RCT are correlated with the variables determining selection. We conclude in Section 6.

2 Related work

There is surprisingly little systematic work studying selection into field experiments. The work most related to ours is the work on selection issues arising in policy evaluation studies (Heckman and Vytlacil (2006)). This work derives theoretical implications regarding the sign of the selection bias one should expect in these interventions. The general prediction is that one should expect a positive selection bias: those who self-select in the intervention are likely to be those who expect to benefit most from the intervention.

Malani (2008) develops a model of self-selection into medical randomised controlled trials where participants self-select into the experiment based on a prior belief about the effectiveness of the treatment. He assumes that patients have a choice between an old treatment and participating in a randomised controlled trial (RCT) involving a new treatment. As the probability of being assigned to the treatment group in the RCT increases, less optimistic patients are willing to participate in the RCT. If the probability is lower than 1 (which will always be the case by definition in a RCT), only the most optimistic will be willing to experiment and we will obtain a positive selection bias. The main difference with our model is that people cannot access the new treatment out of the RCT. While this may be the case in medical trials, it is usually not the case for experiments conducted in the social sciences. This is why those who are optimistic may opt out and prefer to implement the intervention (or a similar intervention) themselves.

In a related paper (Belot and James, 2013), we extend Malani's model to the context of randomised social and policy experiments. A key difference is that in a social or policy experiment, the decision to participate to a randomized controlled trial does not only entail a choice between participating or not doing anything. A third alternative could be to conduct a similar intervention without taking part and without running the risk of being in the control group. This alternative is likely to be attractive for the most optimistic - those who believe the intervention may be effective. They may prefer to opt out from the experiment because there is a chance they will end up in the control group and not receive the benefits from the intervention. In our context, the idea of using incentives in a health-related context is not new. In fact there is a growing market for stickers and other rewards to encourage children to eat healthily. Thus, it is plausible that schools would have wanted to implement an incentive scheme independently of being

approached by us. But of course if they participate and end up in the control group, we will explicitly ask them not to implement any incentive scheme over the course of the study. This example is, we believe, quite representative for many randomised controlled experiments.Thus, two types of selection (positive and negative) could take place at the same time. On the one hand, we have a group of "optimists" (who expect the treatment to be successful) who could opt out and introduce a negative selection bias. On the other hand, we have a group of "pessimists" (who do not expect the treatment to be successful), who could opt out as well if they are not fully compensated for the experimentation and implementation costs. Their decision not to participate will introduce a positive selection bias. Because both types of selection could take place at the same time, it is not possible to sign the direction of the selection bias.

To our knowledge, there is only one other study, by Allcott and Mullainathan (2011), who focus on the issue of external validity and partner selection bias in randomized controlled experiments (RCTs) in the social sciences. They provide evidence of the limited external validity of RCTs with unique data from 14 nearly-identical energy conservation field experiments conducted at different sites in the US. These RCTs are run by a company called OPOWER and the treatment in these experiments is to mail Home Energy Reports (HERs) to residential electricity consumers, with the goal of causing them to use less energy. The fact that the same experiment was conducted at 14 different sites allows them to gauge the extent to which treatment effects vary across sites and thereby provide evidence of the biases that may arise by conducting an experiment in one geographical area alone. The estimated average treatment effects vary by a factor of 2.4, and cost effectiveness by a factor of 4.2. They show that observable population characteristics explain very little of this variation. Relying on a treatment effect estimated in one specific area could therefore lead to very biased estimates of the average treatment effect.

They also discuss various mechanisms of partner selection bias in randomized controlled trials (RCTs) – both positive and negative. There could be a positive bias because implementing randomized trials requires managerial ability and operational efficacy and partners who run the most effective programmes may also be the best equipped to evaluate these programme. This form of positive partner selection bias has been called "gold plating" (Duflo, Glennerster, and Kremer 2007). Another form of positive partner selection bias results from the fact that those running effective programmes are more keen to show they work than those who fear they are running ineffective ones, and prefer to remain "ignorant" (Pritchett 2002). Negative partner selection bias could arise if those who are running the most effective treatments have already treated the parts of their

population that have the largest treatment effects ("diminishing returns bias").

Using data from 939 electricity utilities in the U.S. with more than 10,000 residential customers, they show that those who partner with OPOWER have observable characteristics that are negatively correlated with treatment effects – that is, there is evidence of negative selection. They attribute this negative selection bias to a "diminishing return bias", those with larger treatment effects have already treated their population. They provide further evidence of partner selection bias in micro-finance field experiments and show that micro-financial institutions (MFIs) that have partnered to carry out randomized trials with three large academic initiatives - the Jameel Poverty Action Lab, Innovations for Poverty Action, and the Financial Access Initiative - differ from the average MFI on characteristics that might be associated with treatment effects, including for-profit status, size, experience, share of borrowers that are female, and average loan size.

Only one other paper examines selection in a field experiment setting, Gautier and Van Der Klaauw (2012). Although this does not involve partner selection it shows in a unique and natural setting the role that self-selection can play in leading to biased results in field (and lab) experiments. They are able to examine the effect of selection in a gift exchange game by exploiting a promotion in all hotels of a chain in Belgium and the Netherlands that allowed customers to pay what they wished for a one-night stay. They compare the behaviour of two groups of participants, an involuntary group who had booked before the annoncement of the promotion and were not aware of it at the time of booking with a group who booked after the promotion was announced - a voluntary group. The self-selected voluntary group paid less and did not respond to an exogenous change in the posted price unlike the involuntary group. These results suggest that experiments aiming at testing for pro-social behaviour are underestimating its presence in particular where participants can self-select into the experiment.

To summarise, there are few studies documenting the issue of self selection in randomised controlled experiments in social sciences and, in theory, the sign of the selection bias could go either way.

3 Context and Experimental Design

Obesity and diet are high on the policy agenda. The broad question of our study (which we used to approach the local public authorities), is whether providing short-term incentives can successfully encourage children to make healthier nutritional choices and develop healthier habits. We contacted all 150 local authorities in the United Kingdom on Friday

2 July 2010 and Monday 5 July 2010 (we randomly selected half of the sample to be sent on each day). The e-mail (see appendix) described who we are and the aim of the project. We specifically indicated that we were interested in comparing the effectiveness of incentive schemes to increase the consumption of fruit and vegetables at lunch in schools and that the interventions were designed to target children who have been found to respond little to health interventions, such as children from poorer socio-economic backgrounds and boys. These specific observables (consumption of fruit and vegetables, children from poorer socio-economic backgrounds) vary across local authorities and we are able to document precisely how selection takes place based on the observables we mentioned in the letter.

The timing of this initial contact coincided with the First National Child Obesity Week[2] taking place in the UK, from July 5th until July 10th. School meals and children's diet were also widely discussed in the media. On 30 June 2010, the Secretary of Health Andrew Lansley criticized the Jamie Oliver School Food revolution, referring to the campaign led by the British chef Jamie Oliver aimed at improving parental cooking skills across the country (the criticism was targeted at the patronizing approach of the campaign). The public criticism triggered a strong reaction in the media and the public. Lansley apologised to Oliver later in October 2010. At the same time, the climate regarding public funding was relatively grim. Large public funding cuts were expected across the whole country.

The Experimental Treatments

The experiment is fully documented in Belot, James and Nolen (2013). We will only briefly describe the key elements of the experiment here. The goal of the experiment is to study the effects of providing incentives on choosing and eating healthy items using two different interventions: a piece-rate scheme and competition scheme. In both schemes, pupils were given a sticker for choosing or bringing in a healthy item at lunch. Then, at the end of the week (Friday afternoon), each student had the opportunity to win a larger prize. In the piece-rate scheme, pupils were eligible for an additional reward if they collected four stickers or more during the week. In the competition scheme, students were assigned to random groups of four, and the pupil with the most stickers in each group won an additional reward.

Our initial e-mail mentioned that we intended to test the effectiveness of incentive schemes, without being specific. We also mentioned being particularly interested in studying how boys would respond in comparison to girls, and how pupils from disadvantaged

[2]http://www.mendcentral.org/ncow

backgrounds would respond in comparison to pupils from richer backgrounds. These specific interests came as a follow-up from a previous study we conducted in the UK (Belot and James, 2011).

We now turn to the documentation of the selection in the context of our study.

4 Determinants of Selection

4.1 Selection at the level of local authorities

We first study the determinants of selection at the level of local authorities. We differentiate between response to the e-mail (positive or negative), interest (positive response), collaboration (provision of names of 5 schools in the area), participation (schools in areas decide to take part in the intervention). Out of the 150 local authorities contacted, 63 responded within a month. 33 indicated a positive interest in collaborating and 18 asked for a meeting. 28 agreed to collaborate, and 22 of them have effectively engaged in the first step of collaboration (we asked them to suggest the names of 5 comparable schools in the area). 12 LEAs finally participated in the actual experiment.

4.1.1 Contemporary indicators

We first consider socio-economic and health indicators which are most associated at the local authority level for the year 2008, which were the most recent available indicators in June 2010. The next section discusses the role of pre-intervention trends.

We have information on the average weekly household income, the percentage of free school meals (school meals are part of a means-tested programme and around 17% of children receive free school meals in the UK), the number of schools in the area, and a number of variables capturing overweight children and obesity rates, as well as adult health habits (and in particular, the percentage of people eating the recommended five portions of fruit and vegetables per day).

Regarding the initial response first (Table 1 Panel A) – a response can be either negative (0) or positive (1). Each of our definitions of selection are estimated using a probit regression, and the estimates in each of the tables are presented as marginal effects. We find that areas with a larger number of schools were more likely to respond, and we find some evidence (although the coefficients are not always significant) that authorities with a higher income and a lower consumption of fruit and vegetables were more likely to do so. Turning to the probability of a positive interested response (Table 1 panel B) (and

assuming that no response within a month is a negative response), we find that larger LEAs and LEAs with richer households are more likely to be interested. We find that those with a lower consumption of fruit and vegetables are also more likely to respond, as well as those more confronted with a child obesity problem.

We include additional health indicators to gauge whether the selection is driven by the general state of health in the local area or whether it operates mainly through the indicators and variables that are specifically targeted in the study (children's obesity and diet). Of course these general indicators are correlated with the more specific ones, but the question is whether these general indicators alone are good predictors of interest. In columns (5) and (6) we substitute indicators for smoking and binge drinking with the indicator of fruit and vegetable consumption. We find that these indicators are poor predictors of interest in collaboration – the coefficients are close to zero and rather precisely estimated. This suggests that the interest is not reflecting a general poor state of health in the area but is rather reflecting specific weaknesses with respect to the indicators specifically targeted in the study.

Examining those who collaborate with us, Table 2 panel A, we find less systematic differences in terms of observable characteristics than for the initial interest. The number of schools, and income remain significant. Obesity rates are not systematically different. We also fail to find systematic differences according to the percentage of free school meal children. There remains a systematic difference, albeit to a lesser extent, in terms of fruit and vegetable consumption, which is perhaps the most obvious outcome of interest in our study. Table 2 Panel B shows little selection along the lines of actual participation.

Tables B1-B4 examine determinants that, although not directly associated, could determine whether they respond and express an interest in the intervention. We include the level of education at the age groups we would target in the intervention – key stage 1 (taken at aged 6) and key stage 2 (taken at aged 11). Those with better key stage 2 scores are more likely to respond, however, this is not the case for expressing a positive interest. Given the political timing in which we contacted the local authorities there was a huge amount of uncertainty surrounding funding. We then consider three types of public spending measures. Per pupil spending for the current year, the change in spending, and the spending at the local authority level, none of which determine participation at any level. Finally, we consider the gender of the initial two contacts, the CEO of the local authority and the director of children's services, and the most likely contact to whom our original letter would have been passed on to, the healthy schools contact. The gender of these contacts does not determine response, interest or collaboration.

Summarizing the results, we find some evidence of selection taking place, but it is mild, certainly when it comes to the actual collaboration.

4.1.2 Pre-trends

We now consider the role of pre-trends in health and education indicators. The question is whether the areas that responded and engaged in collaboration with us were on different trends than others. For example, are these areas confronted with a deterioration in these health indicators or, on the contrary, are they the ones on a positive trend, which could indicate that they are effectively engaged in improving children's diet and obesity already?

Table 3 panel A controls for trends in obesity and overweight rates. We find no evidence of any selection based on those trends. Of course it could be that both types are more likely to self-select: more than one standard deviation up or one than more standard deviation down in comparison to 2006 level. Panel B controls for trend in education, again we find no evidence of self-selection.

4.1.3 Non monotonicity

As discussed earlier, a priori there could be positive and negative selection taking place at the same time, that is, both the optimists and the pessimists could choose to opt out. To the extent that these beliefs are correlated with observables in a monotonic way, then one way to allow for positive and negative selection at the same time is to allow for non monotonicity in the relationship between observables and the probability of opting in.

Table 4 replaces the continuous measures of the key variables (fruit and vegetable, obesity, and income) with dummies representing the mid-point and an upper value for that variable. These show that those who self-select are those with the greatest problems to solve. The areas with the highest levels of fruit and vegetable consumption are the least likely to express an interest or collaborate. The areas with the top third of fruit and vegetable consumption are between 32 and 29 percentage points more likely to express and interest, around 25 percentage points to collaborate, however there are no systematic differences in the propensity to collaborate or participate.

4.2 Selection of Schools

We turn to the selection of schools that were recruited into the intervention. As mentioned previously, we asked each local authority to provide the names of 5 schools that would be

willing to participate in the experiment. As we mentioned earlier, the rationale for doing this is to understand how overarching bodies themselves select potential participants for randomised controlled experiments.

We asked for 5 representative schools from the local area, as comparable as possible in terms of socio-economic status and engagement in health interventions, without further requirements. We have quite detailed information about the schools in each local authority. We have information on free school meal participation, school size and a range of school spending variables. In addition to that, we also have information from school inspections conducted by the official inspection body Ofsted[3]. These inspections are summarized in reports that provide measures of how well the school operates or performs in various dimensions. There is a general measure of school performance, which assigns scores from 1 to 4, where a score of 1 is outstanding, 2 is good, 3 is satisfactory, and 4 is inadequate. There is also a measure that provides information on how well the school performs at *"getting its pupils to adopt healthy lifestyles"*, which is obviously directly relevant in this context. Schools are rated according to a similar scoring rule (from 1 to 4) as the general measure. This measure is the closest indicator we have that captures the involvement of schools in promoting healthy lifestyles.

Table 5 presents the results of the analysis of selection. The dependent variable in each case is binary represented by a one if the school was selected to participate by the local authority and zero otherwise. We find a negative selection with regards to an overall indicator of school performance (Ofsted score 1). But we find a positive selection with regard to the health promotion measure. The schools that are selected by the local authority are more likely to be those with an "outstanding" score in promoting healthy lifestyles.

5 Treatment Effects and Selection Characteristics

In this section we examine the treatment effect of the experiment and examine whether this varies according to the characteristics that have been shown to, albeit mildly, determine selection into the experiment. In particular we focus on the immediate effect of the intervention and focus on how the treatment effect differs according to the ofsted health score and the level of fruit and vegetable consumption.

[3]www.ofsted.gov.uk/

In particular we estimate the following:

$$\begin{aligned} Mean(Outcome)_{isz} &= \beta_0 + \beta_z Intervention_z + \beta_{comp} Comp_s + \beta_{piece} PieceRate_s \\ &+ \beta_{z,comp} Comp_s \times Intervention_z \\ &+ \beta_{z,piece} PieceRate_s \times Intervention_z + \alpha_i + u_{ist} \end{aligned} \tag{1}$$

Where the two outcomes we consider are whether the fruit or vegetable was chosen and whether at least some of it was eaten (tried). The mean is the number of items chosen (or tried) during the week by the number of days the student was present. The intervention is a dummy indicating 1 for weeks 2 to 5 and 0 for the baseline week (week 1) and post intervention week (week 6). Comp and PieceRate are dummy variables indicating whether school s is part of the competition or treatment group, and the unobserved error term is $\alpha_i + u_{ist}$

Table 6 presents the estimates from the experiment[4]. Panel A presents the estimates for choice for the whole sample. Column 1 presents the baseline estimates for all pupils. Column 2 shows the effect for those schools not deemed outstanding along the lines of the ofsted health score described in the previous section and column 3 presents the estimates for those which are outstanding. Column 4 shows the p-value of a test of the difference between these two groups of schools. We find no statistically significant difference between the treatment effects for these two groups. Columns 5, 6 and 7 repeat this exercise splitting the sample between those schools in local authorities with below (column 5) and above (column 6) the median of fruit and vegetable consumption at the local authority level for the adult population. The point estimates are larger for both piece rate and competition treatments for the below median fruit and vegetable group, however, they are both imprecisely estimated and are not significantly different from each other.

Panel B considers another group of interest - those who did not choose fruit or vegetable 100% of the time during the baseline week. This group has some margin to improve. There are positive effects for both treatments, albeit imprecisely estimated for the piece rate treatment. We do not find significant differences between the ofsted health groups, nor for above and below the median of fruit and vegetable consumption. Panel C and D

[4]A full account of the treatment effects of the experiment is documented in Belot, James and Nolen (2013). In particular the main estimates examine how the effects vary week by week, this is not our primary interest here and as such we just examine the effect using a dummy variable for the 4 weeks of the intervention period.

present the treatment effects for trying a fruit or vegetable. Again we find no significant differences between the treatment effects for these groups.

6 Discussion and conclusion

This paper discusses and provides evidence of selection of field experimental partners in a policy relevant experiment. We provide evidence based on a specific policy experiment, which consists of providing incentives to children to eat fruit and vegetables at lunch. All potential partners were approached at the same time. We have access to a range of relevant observable measures likely to be correlated with prior beliefs about the effectiveness of the intervention and with the ability to conduct the intervention. We observe mild selection along observable characteristics. Local authorities interested in collaborating tend to be larger and richer, and have lower levels of fruit and vegetable consumption. Next to that, the schools suggested by local authorities and ultimately participating to the experiment tend to be those who are already outstanding in promoting healthy lifestyles. Local authorities are more likely to involve schools that are already engaged in promoting healthy lifestyles, but who do not score highly on an overall evaluation measure of management.

In conclusion, the schools that are part of the experimental sample do not show strong systematic differences along observable socio-economic characteristics, but they are those who have possibly experimented before and are more likely to be located in areas where there is a problem to solve.

Field experimental research has experienced a boom in popularity over the last decade, particularly in the fields of policy evaluation, development economics and personnel economics. Surprisingly, there has not been much attention devoted to the documentation of partner selection. We would argue that a careful documentation of the selection process on observable characteristics would significantly improve the external validation of the results.

References

Allcott, H., and S. Mullainathan (2011), External Validity and Partner Selection Bias,Working Paper, MIT Department of Economics

Altonji, J, T. Elder and C. Taber (2005), Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, *Journal of Political Economy*.

Bandiera, O., I. Barankay and I. Rasul (2007), Incentives for Managers and Inequality Among Workers: Evidence From a Firm-Level Experiment, Quarterly Journal of Economics 122(2), 729-773.

Belot, M. and J. James (2011), Healthy School Meals and Educational Outcomes, Journal of Health Economics 30 (3), 489-504

Belot, M. and J. James (2013), Selection in Field Experiments, mimeo

Belot, M., J. James and P. Nolen (2013), Incentives and Children's Dietary Choices: A Field Experiment in Primary Schools, mimeo

Burtless, G. (1995), The Case for Randomized Field Trials in Economic and Policy Research, *Journal of Economic Perspectives* 9(2), 63-84.

Card, D., S. Dellavigna and U. Malmendier (2011), The Role of Theory in Field Experiments, with David Card and Ulrike Malmendier, Journal of Economic Perspectives, 25(3), 39-62.

Duflo, Esther & Michael Kremer (2005). "Use of Randomization in the Evaluation of Development Effectiveness." in George Pitman, Osvaldo Feinstein and Gregory Ingram (eds.) Evaluating Development Effectiveness. New Brunswick, NJ: Transaction Publishers.

Duflo, E., Dupas, P., & M. Kremer (2011), Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, American Economic Review 101(5), 1739-1774

Dupas, P. (2011), Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya, American Economic Journal: Applied Economics 3(1), 1-34.

Falk, A. (2007), Gift Exchange in the Field, Econometrica 75(5), 1501-1511.

Fehr, E. & L. Goette (2007), Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment, American Economic Review 97(1), 298-317.

Fryer, R., Financial Incentives and Student Achievement: Evidence from Randomized Trials (2011), *Quarterly Journal of Economics*, 126(4): 1755-1798

Gautier, P. and B. van der Klaauw (2012), Selection in a field experiment with voluntary participation, *Journal of Applied Econometrics*, 27, 63-84.

Hastings, J. and J. Weinstein, School Choice, and Academic Achievement: Evidence from Two Experiments (2008), *Quarterly Journal of Economics*, 123(4), 1373-1414

Harrison, Glenn, W., and John A. List, 2004. Field experiments. *Journal of Economic Literature* 42, 1009105

Heckman, J.J. and E. Vytlacil (2001), Policy-Relevant Treatment Effects, *The American Economic Review* 91(2), 107-111

Heckman, James and E.J. Vytlacil (2006), Econometric Evaluation of Social Programs. *Handbook of Econometrics*, Volume 6, James Heckman and E. Leamer, eds., Amsterdam: North Holland.

Heckman, James, R. Lalonde and J. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier Science.

Holt, Charles, A. (2005), *Markets, Games and Strategic Behavior: Recipes for Interactive Learning*. Addison-Wesley, Boston

List, John (2011), Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives*, 25(3): 316.

Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives*, 25(3): 1738.

Oster, E., and R., Thornton, (2011) Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation, American Economic Journal: Applied Economics 3(1): 91-100

Wisdom, J., J.S. Downs and G. Loewenstein (2010), Promoting Healthy Choices: Information versus Convenience, American Economic Journal: Applied Economics, 2(2), 164-78.

Figures and Tables

Figure 1

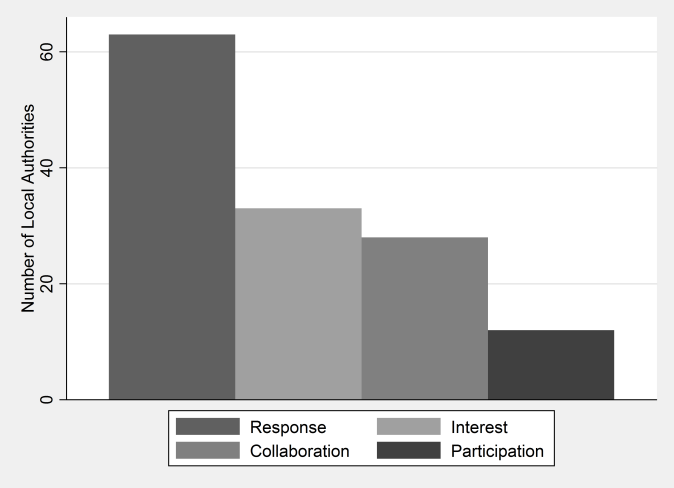

Table 1: Main Determinants of response to initial e-mail, interest and response

	Panel A: Response					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.142	0.147*	0.141	0.140	0.124	0.123
	(0.087)	(0.087)	(0.087)	(0.087)	(0.086)	(0.087)
Contacted by J James	0.076	0.073	0.075	0.072	0.085	0.085
	(0.085)	(0.085)	(0.085)	(0.085)	(0.084)	(0.084)
Income/100	0.084*	0.082	0.084*	0.085*	0.016	0.007
	(0.050)	(0.050)	(0.051)	(0.050)	(0.038)	(0.038)
% FSM	0.741	0.129	0.660	0.344	0.759	0.843
	(0.570)	(0.727)	(0.782)	(0.808)	(0.591)	(0.525)
Number of schools/100	0.080**	0.080**	0.080**	0.081**	0.071*	0.070*
	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)
Fruit and Veg	-0.028*	-0.026*	-0.028*	-0.027*		
	(0.015)	(0.015)	(0.016)	(0.015)		
% Overweight & Obese Reception	-0.000					
	(0.019)					
% Obese Reception		0.044				
		(0.036)				
% Overweight & Obese Year 6			0.003			
			(0.021)			
% Obese Year 6				0.016		
				(0.025)		
Smoking					0.004	
					(0.014)	
Binge Drinking						-0.001
						(0.012)
Observations	145	145	145	145	145	145
R squared	0.0656	0.0731	0.0657	0.0676	0.0482	0.0479
	Panel B: Interest					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.066	0.070	0.043	0.044	0.046	0.052
	(0.080)	(0.080)	(0.081)	(0.081)	(0.080)	(0.082)
Contacted by J James	0.041	0.030	0.034	0.027	0.056	0.053
	(0.077)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)
Income/100	0.140***	0.134***	0.133***	0.143***	0.057*	0.035
	(0.047)	(0.047)	(0.048)	(0.048)	(0.034)	(0.034)
% FSM	0.568	-0.057	-0.122	-0.380	0.566	0.813*
	(0.479)	(0.610)	(0.655)	(0.717)	(0.497)	(0.458)
Number of schools/100	0.087***	0.089***	0.090***	0.095***	0.074**	0.069**
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Fruit and Veg	-0.036**	-0.035**	-0.033**	-0.035**		
	(0.015)	(0.015)	(0.015)	(0.015)		
% Overweight & Obese Reception	0.017					
	(0.018)					
% Obese Reception		0.061*				
		(0.033)				
% Overweight & Obese Year 6			0.035*			
			(0.020)			
% Obese Year 6				0.051**		
				(0.026)		
Smoking					0.015	
					(0.012)	
Binge Drinking						0.003
						(0.011)
Observations	128	128	128	128	128	128
R squared	0.113	0.131	0.128	0.133	0.0657	0.0565

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Main Determinants of response to initial e-mail, collaboration and participation

	Panel A: Collaboration					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.004	0.007	-0.007	-0.006	-0.010	0.000
	(0.067)	(0.066)	(0.067)	(0.066)	(0.068)	(0.068)
Contacted by J James	0.014	0.011	0.010	0.008	0.022	0.021
	(0.064)	(0.064)	(0.064)	(0.064)	(0.065)	(0.065)
Income/100	0.086**	0.083**	0.081**	0.086**	0.030	0.030
	(0.040)	(0.039)	(0.040)	(0.040)	(0.029)	(0.029)
% FSM	0.425	0.055	0.092	-0.141	0.437	0.580
	(0.408)	(0.517)	(0.563)	(0.595)	(0.425)	(0.382)
Number of schools/100	0.071***	0.072***	0.072***	0.075***	0.063**	0.062**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Fruit and Veg	-0.024*	-0.024*	-0.022*	-0.023*		
	(0.012)	(0.012)	(0.012)	(0.012)		
% Overweight & Obese Reception	0.009					
	(0.015)					
% Obese Reception		0.035				
		(0.026)				
% Overweight & Obese Year 6			0.017			
			(0.016)			
% Obese Year 6				0.028		
				(0.019)		
Smoking					0.008	
					(0.010)	
Binge Drinking						0.007
						(0.009)
Observations	145	145	145	145	145	145
R squared	0.0748	0.0840	0.0792	0.0864	0.0460	0.0458
	Panel B: Participation					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.062	0.064*	0.058	0.061	0.060	0.064
	(0.040)	(0.038)	(0.040)	(0.040)	(0.040)	(0.042)
Contacted by J James	0.014	0.012	0.011	0.012	0.013	0.012
	(0.032)	(0.028)	(0.033)	(0.033)	(0.033)	(0.033)
Income/100	-0.029	-0.030	-0.033	-0.031	-0.029	-0.019
	(0.022)	(0.020)	(0.023)	(0.023)	(0.019)	(0.021)
% FSM	-0.402	-0.583**	-0.520	-0.435	-0.448	-0.396
	(0.282)	(0.288)	(0.354)	(0.386)	(0.322)	(0.279)
Number of schools/100	0.015	0.015	0.015	0.015	0.015	0.015
	(0.013)	(0.012)	(0.013)	(0.014)	(0.013)	(0.013)
Fruit and Veg	-0.000	-0.001	-0.000	-0.002		
	(0.006)	(0.005)	(0.007)	(0.007)		
% Overweight & Obese Reception	0.009					
	(0.008)					
% Obese Reception		0.024*				
		(0.013)				
% Overweight & Obese Year 6			0.007			
			(0.009)			
% Obese Year 6				0.003		
				(0.011)		
Smoking					0.004	
					(0.005)	
Binge Drinking						0.007
						(0.006)
Observations	145	145	145	145	145	145
R squared	0.159	0.185	0.150	0.144	0.148	0.160

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3: Role of Pre-Trends in Obesity and Education

	Panel A: Obesity Trends							
	Response		Interest		Collaboration		Participation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted by J James	0.090	0.091	0.059	0.060	0.020	0.021	0.014	0.015
	(0.084)	(0.084)	(0.076)	(0.076)	(0.064)	(0.064)	(0.035)	(0.035)
Income/100	0.076	0.073	0.129***	0.126***	0.081**	0.079**	-0.034	-0.032
	(0.050)	(0.050)	(0.047)	(0.048)	(0.039)	(0.040)	(0.024)	(0.024)
% FSM	0.823	0.636	0.739	0.561	0.514	0.370	-0.269	-0.180
	(0.588)	(0.737)	(0.511)	(0.644)	(0.440)	(0.538)	(0.333)	(0.381)
Number of schools/100	0.093**	0.094**	0.091***	0.093***	0.071***	0.072***	0.022	0.022
	(0.038)	(0.038)	(0.032)	(0.032)	(0.027)	(0.027)	(0.014)	(0.014)
Fruit and Veg	-0.025	-0.025	-0.037**	-0.037**	-0.024**	-0.025**	-0.002	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)	(0.012)	(0.012)	(0.007)	(0.007)
Obesity 2008-Obesity 2006 (Reception)	-0.005		-0.005		0.000		-0.000	
	(0.022)		(0.020)		(0.017)		(0.009)	
Obesity 2008-Obesity 2006 (y6)		-0.009		-0.009		-0.005		0.003
		(0.019)		(0.018)		(0.015)		(0.008)
Observations	146	146	129	129	146	146	146	146
R squared	0.0508	0.0517	0.0991	0.100	0.0702	0.0711	0.111	0.113
	Panel B: Education Trends							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted by J James	0.092	0.089	0.034	0.030	0.011	0.008	0.017	0.016
	(0.085)	(0.084)	(0.077)	(0.077)	(0.063)	(0.063)	(0.032)	(0.032)
Income/100	0.078	0.074	0.137***	0.139***	0.081**	0.085**	-0.032	-0.034
	(0.050)	(0.050)	(0.048)	(0.048)	(0.040)	(0.039)	(0.022)	(0.022)
% FSM	0.454	0.313	-0.405	-0.252	-0.250	-0.108	-0.256	-0.309
	(0.843)	(0.855)	(0.733)	(0.755)	(0.613)	(0.612)	(0.367)	(0.384)
Number of schools/100	0.094**	0.092**	0.098***	0.097***	0.073***	0.073***	0.026*	0.024*
	(0.038)	(0.038)	(0.032)	(0.031)	(0.026)	(0.026)	(0.014)	(0.013)
% Obese Reception	0.040	0.040	0.030	0.032	0.020	0.019	0.025	0.026
	(0.043)	(0.043)	(0.040)	(0.039)	(0.032)	(0.032)	(0.017)	(0.017)
% Obese Year 6	-0.001	-0.000	0.037	0.036	0.019	0.021	-0.006	-0.007
	(0.030)	(0.030)	(0.031)	(0.031)	(0.023)	(0.024)	(0.012)	(0.012)
KS1 Score 2009-KS1 Score 2007	-0.155		-0.052		0.055		-0.085	
	(0.226)		(0.199)		(0.165)		(0.090)	
Fruit and Veg	-0.022	-0.022	-0.032**	-0.033**	-0.022*	-0.022*	-0.001	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)	(0.012)	(0.012)	(0.006)	(0.006)
KS2 Score 2009-KS2 Score 2007		0.031		-0.206		-0.154		-0.038
		(0.243)		(0.228)		(0.184)		(0.101)
Observations	146	146	129	129	146	146	146	146
R squared	0.0592	0.0569	0.131	0.136	0.0870	0.0912	0.155	0.145

notes: All coefficients presented as marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 4: Splines of key variables

	Response		Interest		Collaboration		Participation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted on Friday	0.125	0.120	0.039	0.044	-0.009	-0.008	0.017	0.011
	(0.087)	(0.088)	(0.080)	(0.080)	(0.065)	(0.065)	(0.020)	(0.015)
Contacted by J James	0.099	0.116	0.069	0.074	0.023	0.034	-0.001	0.003
	(0.085)	(0.086)	(0.077)	(0.078)	(0.063)	(0.063)	(0.010)	(0.007)
% FSM	1.137	0.210	0.476	0.009	0.305	-0.021	-0.076	-0.120
	(0.728)	(0.651)	(0.633)	(0.554)	(0.513)	(0.447)	(0.108)	(0.129)
Number of schools/100	0.072*	0.084**	0.083**	0.086**	0.066**	0.071**	0.000	0.001
	(0.040)	(0.040)	(0.034)	(0.034)	(0.028)	(0.028)	(0.004)	(0.003)
Income/100	-0.003	0.018	0.082	0.077	0.053	0.055	-0.049	-0.035
	(0.064)	(0.065)	(0.056)	(0.057)	(0.047)	(0.047)	(0.037)	(0.033)
Income Mid	0.180	0.133	0.162	0.104	0.107	0.067	0.049	0.030
	(0.128)	(0.131)	(0.129)	(0.130)	(0.108)	(0.106)	(0.049)	(0.036)
Income Upper	0.228	0.106	0.205	0.110	0.170	0.095	0.616**	0.508
	(0.216)	(0.227)	(0.215)	(0.216)	(0.189)	(0.182)	(0.283)	(0.316)
Fruit & Veg Mid	-0.118	-0.084	-0.099	-0.069	-0.086	-0.070	0.018	0.023
	(0.120)	(0.122)	(0.100)	(0.106)	(0.079)	(0.081)	(0.026)	(0.029)
Fruit & Veg Upper	-0.210	-0.174	-0.323***	-0.287**	-0.266***	-0.246***	-0.003	0.001
	(0.148)	(0.149)	(0.110)	(0.116)	(0.092)	(0.094)	(0.019)	(0.015)
Obesity Mid (y6)	-0.010		0.233*		0.107		-0.007	
	(0.117)		(0.125)		(0.102)		(0.011)	
Obesity Upper (y6)	-0.084		0.158		0.093		-0.016	
	(0.152)		(0.168)		(0.135)		(0.021)	
Obese Mid (Reception)		0.023		0.026		0.005		-0.008
		(0.116)		(0.112)		(0.088)		(0.011)
Obese Upper (Reception)		0.236*		0.286*		0.204		0.009
		(0.142)		(0.157)		(0.135)		(0.018)
Observations	147	147	130	130	147	147	147	147
R squared	0.0678	0.0831	0.144	0.152	0.109	0.129	0.289	0.318

notes: a) All coefficients presented as marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1 b) Mid refers to above the 33rd and below (and equal to) the 66th centiles, upper refers to above the 66th percentile, the reference group is below (or equal to) the 33rd pecentile.

Table 5: Selected Schools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ofsted Health Score 1	0.017**	0.020**	0.021**	0.020**	0.017**	0.018**	0.014*
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.007)
Ofsted Health Score 3 and 4	0.005	0.012	0.009	0.007	0.007	0.007	0.007
	(0.010)	(0.015)	(0.014)	(0.013)	(0.011)	(0.011)	(0.011)
Ofsted Overall Score 1		-0.017***	-0.016***	-0.015***	-0.012***	-0.015***	-0.012***
		(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)
Ofsted Overall Score 3 and 4		-0.008	-0.008	-0.009	-0.006	-0.005	-0.006
		(0.008)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)
FSM Medium			0.012	0.010	0.029	0.029	0.025
			(0.016)	(0.015)	(0.024)	(0.024)	(0.023)
FSM Low			-0.006	-0.005	0.009	0.008	0.007
			(0.012)	(0.011)	(0.008)	(0.009)	(0.008)
Total Pupils/100				0.005***	0.004**	0.005**	0.002
				(0.002)	(0.002)	(0.002)	(0.002)
Catering Staff Costs/100					0.011***	0.009*	0.004
					(0.004)	(0.005)	(0.006)
Other Staff Costs					-0.000	-0.000	-0.000
					(0.000)	(0.000)	(0.000)
Energy Costs					0.015**	0.013*	0.010
					(0.007)	(0.007)	(0.007)
Total School Income/100					-0.001*	-0.000	-0.001
					(0.000)	(0.000)	(0.000)
Avg Eng/Math Score 09/10						0.001	-0.000
						(0.002)	(0.002)
Observations	1,973	1,971	1,971	1,971	1,971	1,789	1,789
R squared	0.0128	0.0292	0.0423	0.0613	0.0981	0.103	0.160
LEA Dummies	No	No	No	No	No	No	Yes

notes: a) All coefficients presented as marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. b) Sample of schools based on paricipating local authorities. c) Ofsted Scores: 1 is outstanding, 2 is good (omitted catergory), 3 and 4 represent satisfactory and inadequate. (source: www.ofsted.gov.uk) d) FSM Band - The three broad bands used to group pupils eligible for FSM are: Low: less than 20%, Medium: 20.01-35% and High: greater than 35% (omitted catergory). This data comes from school level expenditure per pupils tables (source: https://media.education.gov.uk/schoolsspenddata/primary.xls). e) Source of Cost (spending) variables: https://media.education.gov.uk/schoolsspenddata/primary.xls

Table 6: Treatment Effects and Selection Characteristics

	(1) Baseline	(2) Ofsted Health Not Outstanding	(3) Ofsted Health Outstanding	(4) P-value of diff	(5) Below FV Median	(6) Above FV Median	(7) P-value of diff
Panel A: Choice							
Comp*Intervention	0.048	0.034	0.018	0.218	0.076	0.010	0.185
	(0.034)	(0.046)	(0.038)		(0.046)	(0.039)	
Piece-Rate*Intervention	0.047	0.033	0.044	0.911	0.092	0.003	0.886
	(0.055)	(0.090)	(0.068)		(0.079)	(0.073)	
Observations	2,768	1,272	1,496	2,768	1,552	1,216	2,768
Number of pupils	626	295	331	626	351	275	626
Panel B: Less than 100% Choice in baseline week							
Comp*Intervention	0.157***	0.141**	0.103**	0.122	0.170***	0.190*	0.607
	(0.056)	(0.058)	(0.048)		(0.052)	(0.114)	
Piece-Rate*Intervention	0.102	0.097	0.099	0.722	0.125	0.127	0.375
	(0.074)	(0.104)	(0.095)		(0.080)	(0.139)	
Observations	971	473	498	971	601	370	971
Number of pupils	203	95	108	203	121	82	203
Panel C: Try							
Comp*Intervention	0.081**	0.089*	0.051	0.695	0.043	0.118	0.828
	(0.039)	(0.051)	(0.045)		(0.040)	(0.081)	
Piece-Rate*Intervention	0.069	0.058	0.069	0.925	0.099	0.034	0.733
	(0.064)	(0.085)	(0.086)		(0.115)	(0.058)	
Observations	2,679	1,194	1,485	2,679	1,459	1,220	2,679
Number of pupils	609	280	329	609	334	275	609
Panel D: Less than 100% Try in baseline week							
Comp*Intervention	0.181***	0.185***	0.161***	0.407	0.144***	0.179***	0.183
	(0.037)	(0.051)	(0.038)		(0.041)	(0.059)	
Piece-Rate*Intervention	0.165*	0.112	0.180	0.686	0.201	0.103	0.554
	(0.093)	(0.131)	(0.123)		(0.142)	(0.091)	
Observations	1,339	616	723	1,339	671	668	1,339
Number of pupils	304	146	158	304	156	148	304

notes: OLS regression results; including pupil random effects and with standard errors clustered at the school level. Additional controls: week dummies. Dependent variable is the pupil weekly mean. *, ** and *** indicate significance at 10%, 5% and 1% level respectively. Column (5) restricts the sample to schools where the fruit and vegetable consumption at the local authority level is equal to or below 26.5% and column (6) restricts the sample to schools where the fruit and vegetable consumption is above 26.5%

Appendix A: Survey of Top 5 field experimental papers

Policy (health and education)

Fryer, QJE (2011) Financial Incentives and Student Achievement : Evidence from Randomized Trials

- Partner: 200 schools across three cities in the US
- Intervention: Provision of incentives to school children (on school inputs and outputs)
- Documentation on partner selection: Yes
- Comparison with broader sample of the population: Yes

Hastings and Weinstein, QJE (2008)

- Partner : Charlotte Mecklenburg Public School Distict
- Intervention: Provision of information about school outcomes to parents
- Documentation on partner selection: No
- Comparison with broader sample of the population: No

Wisdom, Downs, and Loewenstein, AEJ:Applied (2010)

- Partner: Fast Food Sandwich Chain
- Intervention: Provision of nutritional information
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Development Economics

Dupas, AEJ:Applied (2011)

- Partner: two rural districts of Western Kenya, involving 328 primary schools
- Intervention: Provision of HIV risk information to teenagers
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Duflo, Dupas and Kremer, AER (2011)

- Partner: Schools in Kenya
- Intervention: To examine the impact of tracking in schools
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Oster and Thornton, AEJ:Applied (2009)

- Partner: Four schools in Chitwan District, Nepal
- Intervention: Test the impact of providing sanitary products of school attendance
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Personnel economics

Bandiera, Barankay and Rasul, QJE (2007)

- Partner: fruit picking farm
- Goal: Test the effectiveness of worker's pay incentive scheme
- Documentation on partner selection: No
- Comparison with broader sample of the population: No

Falk, Econometrica (2007): Gift exchange in the field,

- Partner: Charitable organization
- Goal: Testing gift exchange in the field
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Fehr and and Goette, AER (2007)

- Partner: Bicycle Messenger Service
- Goal: Testing labour supply responses to transitory wage changes
- Documentation of partner selection: No
- Comparison with broader sample of the population: No

Appendix B: Additional Tables

Table B1: Other Determinants of response to initial e-mail, Response

	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.147*	0.147*	0.141	0.150*	0.155*	0.164*
	(0.087)	(0.088)	(0.088)	(0.090)	(0.088)	(0.088)
Contacted by J James	0.073	0.073	0.076	0.064	0.071	0.078
	(0.086)	(0.086)	(0.086)	(0.087)	(0.086)	(0.086)
Income/100	0.073	0.044	0.045	0.079	0.087*	0.093*
	(0.052)	(0.055)	(0.063)	(0.051)	(0.051)	(0.051)
% FSM	0.406	0.409	-0.523	0.179	0.478	0.354
	(0.904)	(0.848)	(1.611)	(0.838)	(0.926)	(0.868)
Number of schools/100	0.080**	0.087**	0.076*	0.075*	0.073*	0.079*
	(0.039)	(0.040)	(0.040)	(0.039)	(0.040)	(0.041)
% Obese Year 6	0.002	-0.005	-0.004	0.002	0.006	0.001
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
% Obese Reception	0.049	0.074	0.034	0.040	0.045	0.050
	(0.044)	(0.047)	(0.046)	(0.045)	(0.044)	(0.044)
Fruit and Veg	-0.026*	-0.023	-0.033*	-0.024	-0.031*	-0.030*
	(0.016)	(0.016)	(0.017)	(0.016)	(0.017)	(0.016)
Key Stage 1: Avg score 2009	0.101					
	(0.140)					
Key Stage 2: Avg score 2009		0.232*				
		(0.130)				
Per Pupil Spending 2010/11			0.000			
			(0.000)			
% Per Pupil Spend increase 2010/11			1.655			
			(28.800)			
% LA Spending Change 2010/11			1.601			
			(3.416)			
Gender CEO				0.002		
				(0.104)		
Gender Children Services				0.074		
				(0.088)		
Gender Healthy Schools				0.015		
				(0.113)		
% of Labour Councillors					-0.349	
					(0.355)	
% of Conservative Councillors					-0.077	
					(0.297)	
Labour Controlled Council						-0.159
						(0.123)
Conservative Controlled Council						-0.035
						(0.118)
Observations	145	145	145	143	145	145
R squared	0.0757	0.0892	0.0794	0.0746	0.0785	0.0810

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

Table B2: Other Determinants of response to initial e-mail, Interest

	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.049	0.050	0.047	0.085	0.051	0.051
	(0.082)	(0.081)	(0.082)	(0.083)	(0.081)	(0.082)
Contacted by J James	0.022	0.021	0.018	0.012	0.019	0.019
	(0.078)	(0.077)	(0.077)	(0.078)	(0.078)	(0.078)
Income/100	0.128**	0.121**	0.129**	0.142***	0.132***	0.134***
	(0.050)	(0.051)	(0.057)	(0.049)	(0.049)	(0.049)
% FSM	-0.237	-0.365	-1.376	-0.447	-0.307	-0.284
	(0.787)	(0.722)	(1.433)	(0.708)	(0.789)	(0.742)
Number of schools/100	0.091***	0.096***	0.097***	0.092***	0.089***	0.084**
	(0.032)	(0.032)	(0.033)	(0.033)	(0.034)	(0.033)
% Obese Year 6	0.037	0.032	0.035	0.040	0.031	0.036
	(0.032)	(0.032)	(0.031)	(0.031)	(0.033)	(0.031)
% Obese Reception	0.041	0.052	0.030	0.033	0.036	0.037
	(0.040)	(0.043)	(0.041)	(0.040)	(0.040)	(0.040)
Fruit and Veg	-0.033**	-0.032**	-0.036**	-0.033**	-0.035**	-0.037**
	(0.015)	(0.015)	(0.016)	(0.015)	(0.016)	(0.015)
Key Stage 1: Avg score 2009	0.101					
	(0.130)					
Key Stage 2: Avg score 2009		0.122				
		(0.117)				
Per Pupil Spending 2010/11			0.000			
			(0.000)			
% Per Pupil Spend increase 2010/11			-6.401			
			(24.347)			
% LA Spending Change 2010/11			-1.011			
			(2.998)			
Gender CEO				0.151		
				(0.103)		
Gender Children Services				0.029		
				(0.079)		
Gender Healthy Schools				-0.037		
				(0.109)		
% of Labour Councillors					0.081	
					(0.317)	
% of Conservative Councillors					0.180	
					(0.282)	
Labour Controlled Council						-0.023
						(0.114)
Conservative Controlled Council						0.107
						(0.119)
Observations	128	128	128	126	128	128
R squared	0.143	0.146	0.143	0.157	0.142	0.147

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

Table B3: Other Determinants of response to initial e-mail, Collaboration

	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	-0.004	-0.002	0.002	0.024	-0.002	0.006
	(0.067)	(0.067)	(0.067)	(0.068)	(0.067)	(0.067)
Contacted by J James	0.006	0.007	0.006	-0.002	0.005	0.010
	(0.064)	(0.064)	(0.063)	(0.063)	(0.064)	(0.064)
Income/100	0.069*	0.072*	0.071	0.084**	0.080**	0.090**
	(0.041)	(0.043)	(0.048)	(0.040)	(0.040)	(0.040)
% FSM	0.134	-0.135	-1.153	-0.178	-0.079	-0.043
	(0.646)	(0.603)	(1.143)	(0.591)	(0.655)	(0.616)
Number of schools/100	0.072***	0.076***	0.080***	0.070**	0.070**	0.069**
	(0.027)	(0.027)	(0.028)	(0.027)	(0.028)	(0.028)
% Obese Year 6	0.026	0.018	0.018	0.022	0.018	0.022
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.024)
% Obese Reception	0.027	0.031	0.015	0.018	0.019	0.022
	(0.032)	(0.035)	(0.033)	(0.032)	(0.032)	(0.032)
Fruit and Veg	-0.022*	-0.021*	-0.021	-0.022*	-0.024*	-0.026**
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Key Stage 1: Avg score 2009	0.159					
	(0.106)					
Key Stage 2: Avg score 2009		0.085				
		(0.099)				
Per Pupil Spending 2010/11			0.000			
			(0.000)			
% Per Pupil Spend increase 2010/11			6.949			
			(19.917)			
% LA Spending Change 2010/11			-1.917			
			(2.548)			
Gender CEO				0.083		
				(0.086)		
Gender Children Services				0.008		
				(0.066)		
Gender Healthy Schools				-0.001		
				(0.085)		
% of Labour Councillors					0.087	
					(0.267)	
% of Conservative Councillors					0.168	
					(0.233)	
Labour Controlled Council						-0.085
						(0.082)
Conservative Controlled Council						0.030
						(0.096)
Observations	145	145	145	143	145	145
R squared	0.105	0.0940	0.0973	0.0937	0.0926	0.0985

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

Table B4: Other Determinants of response to initial e-mail, Participation

	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.061	0.064*	0.062*	0.057	0.064*	0.064*
	(0.038)	(0.038)	(0.037)	(0.038)	(0.037)	(0.038)
Contacted by J James	0.011	0.014	0.009	0.013	0.008	0.010
	(0.026)	(0.028)	(0.024)	(0.026)	(0.025)	(0.025)
Income/100	-0.031	-0.033	-0.032	-0.028	-0.027	-0.024
	(0.020)	(0.021)	(0.022)	(0.019)	(0.019)	(0.019)
% FSM	-0.364	-0.394	-0.919	-0.430	-0.599	-0.541*
	(0.335)	(0.330)	(0.575)	(0.308)	(0.364)	(0.328)
Number of schools/100	0.014	0.015	0.017	0.016	0.016	0.017
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
% Obese Year 6	-0.007	-0.009	-0.008	-0.006	-0.009	-0.007
	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
% Obese Reception	0.029*	0.033*	0.022	0.025	0.026*	0.026*
	(0.016)	(0.018)	(0.016)	(0.016)	(0.015)	(0.015)
Fruit and Veg	-0.002	-0.001	-0.001	-0.001	-0.000	-0.001
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Key Stage 1: Avg score 2009	0.028					
	(0.044)					
Key Stage 2: Avg score 2009		0.029				
		(0.044)				
Per Pupil Spending 2010/11			0.000			
			(0.000)			
% Per Pupil Spend increase 2010/11			10.924			
			(9.506)			
% LA Spending Change 2010/11			-1.123			
			(1.074)			
Gender CEO				0.005		
				(0.035)		
Gender Children Services				0.022		
				(0.029)		
Gender Healthy Schools				-0.037		
				(0.050)		
% of Labour Councillors					0.062	
					(0.108)	
% of Conservative Councillors					-0.043	
					(0.106)	
Labour Controlled Council						-0.005
						(0.036)
Conservative Controlled Council						-0.041
						(0.037)
Observations	145	145	145	143	145	145
R squared	0.198	0.198	0.233	0.210	0.205	0.209

notes:
All coefficients presented as marginal effects from a probit regression, standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1

Appendix C: E-mail sent to local authorities

Dear Mr xxxxxx,

We are researchers in economics at the University of Oxford and the University of Essex and take the liberty to contact you to enquire whether you would be willing to collaborate with us in the context of a research project on children's diet habits.

To give a brief background, we are specifically interested in the effects of diet on health outcomes and cognitive development (educational performance). We have recently conducted a detailed analysis of the effects of the Jamie Oliver Feed Me Better Campaign on educational achievements of young children and found that improving school meals did lead to a significant improvement in Key Stage 2 educational scores and to a reduction in absenteeism. One of the most striking findings we highlighted was the differences across subgroups of children in how marked these improvements were: we found stronger improvements in educational achievements for girls than boys, and we also found that it took longer to see improvements among children from poorer socioeconomic background than among children from more favourable socioeconomic backgrounds. This evidence is in line with findings of other studies. This fact is obviously a serious concern for policy, particularly in the context of reducing disparities in society.

Our ambition would be to assess the effectiveness of interventions targeted at changing diet habits, and specifically those of children from poorer socio-economic backgrounds. Our research expertise is specifically in the design and evaluation of policies using state-of-the art statistical techniques. The interventions we propose are inspired by recent research published in high standards scientific journals in economics and psychology, which suggest that introducing temporary incentives to encourage healthy behaviour might effectively result in long lasting changes in habits.

In a nutshell, the idea would be to conduct interventions in schools in a systematic manner to shed light on the most effective mechanisms to trigger changes in diet habits. We attach a brief description of our proposal to this e-mail.

Given the large interest in public policy for improving diet and health outcomes, we were hoping that you might be interested in collaborating with us. We would be more than happy to provide more details about what we would like to do. You can find all our contact details in the attached proposal.

We very much look forward to hearing from you,

Thanking you very much in advance

Michele Belot and Jonathan James